JCSE International Journal of Computer Sciences & Engineering Open Access
Survey Paper Volume-4, Special Issue-3, May 2016 E-ISSN: 2347-2693

A Survey on the Small Data

Smitha K.A[1], Sangeetha B[2], Ramu B.S[3], D.V Ashoka[4]
[1,2,3,4]*Dept. of Information Science and Engineering,*
JSS academy of Technical Education, Bengaluru, India

Available online at: www.ijcseonline.org

***Abstract*--** Small data deals with data that is small in size for human comprehension. About one quarter of the human brain is involved in visual processing and the only way to comprehend "big data" is to reduce them to small, this data can be informative, accessible and actionable. Small data typically provides information that answers a specific question or addresses a specific problem. This paper highlights the concept of small data and explains how to overcome the difficulties of big data. Small data is about placing the small datasets where it is actually needed. By considering the data of any size solves a specific problem, instead of the massive data which does not solve the problem.

***Keywords*--**Big data, Geographical Information System (GIS), Internet of Things (IoT), FOG computation.

I. INTRODUCTION

From past one-decade people in IT focussing more on Big data. They used this for many applications. Now, they are feeling some complexity involved in big data and it give trouble in a very big manner! To overcome this, software engineers now a day's using the classical way of thinking about data, that is Small data.

Small data build's fundamental level for future that will lead to invest our master level foundation. Small data is for people and big data is for machine [1]. It is true that, one can play with it and analyse and decide the next step with its analysis. It helps in making better decisions.

The small data refers to a combination of structured and unstructured data that may be measured in petabytes or Exabyte. It requires no great shift in technical expertise or any understanding of complex statistics.

"Small data are derived from our individual digital traces". Small data could easily catch the marketers via feedback. They gain control and immediate feedback too from sales campaign that big data systems, still largely residing in the IT departments are unable to duplicate [2].

Small data is small set of specific attributes of Internet of Things. It might be sensor data such as temperature, wind speed and status. Small data-low volumes, batch velocities and structured varieties.

The mass democratization of the means of access, storage and processing of data in IT environment is a complex task. This story is not about large organizations running parallel software on tens of thousands of servers, but about more people than ever being able to collaborate effectively around a distributed ecosystem of information, an ecosystem of small data.

Further because of Small data in many of the data based companies a lot of marketing revolution is in progress. Marketing icons like Allen Bonde, Martin Lindstrom and others suggest that, the small clues serves for greater decision in marketing. Some of the examples are Baseball scores, inventory reports, driving records, sales data, biometric measurements, search histories, weather forecasts and usage alerts, GIS.

II. ICONS ASPECTS

A formal definition of small data has been proposed by "Allen Bonde", Vice President of Innovation at Actuate:

"Small data connects people with timely, meaningful insights (derived from big data and/or "local" sources), organized and packaged – often visually – to be accessible, understandable, and actionable for everyday tasks" [3].

According to Martin Lindstrom:

"Seemingly insignificant behavioural observations containing very specific attributes pointing towards an unmet customer need. Small data is the foundation for break through ideas or completely new ways to turn around brands" [4].

III. ESSENCE OF SMALL DATA

The necessity of small data is as follows [5],

- *Small data at ease:* Marketers and online strategists don't need full-on big data to deliver personalized experiences.
- *Small data is transparent:* One such example is Social channels, we are constantly *creating* this small data each time we check in, search, browse, post etc.
- *Small data ruling Customer Relationship Management (CRM):* Social CRM used to create a complete picture of customers. Small data is the key for building rich profiles of customers which include views from social channels and campaigns with Web analytics and transactional data.
- *Return on Investment (ROI):* Small data can be analysed, actioned and it will be in human

comprehension. Better decisions by proper analysis, assured returns can be expected.

- *Small data is the new revolution in data-driven marketing:* Small data-driven marketing replaces the way businesses interact with customers, transform how customers access and digest useful data, and ultimately redefine the relationship between buyers and sellers.
- *Wearable data of customer*: Small data allows the consumers to modernise their shopping, power their fitness routine, or deliver recommendations about the best price for their next flight. Smart wearable data-driven devices are more market demand for packaged data and data-delivery devices that fill the needs of everyday consumers.
- *User first:* Small data is about the end-user, what they need, and how they can take action. Focus on the user first.
- *Human understandable:* Some small data will start life as big data, to understand one need not be a data scientist to understand or implement it on everyday tasks.

IV. SMALL DATA AND GEOGRAPHICAL INFORMATION SYSTEM

A Geographic Information System (GIS) is a computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface. GIS can show many different kinds of data on one map. This enables people to more easily see, analyse, and understand patterns and relationships.

GIS is a broad term that can refer to a number of different technologies, processes, and methods. It is attached to many operations and has many applications related to engineering, planning, management, transport/logistics, insurance, telecommunications, and business. For that reason, GIS and location intelligence applications can be the foundation for many location-enabled services that rely on analysis and visualization.

GIS has been using big data, wherein big data is the term for a collection of data sets which are so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges of big data include capture, storage, search, sharing, transfer, analysis and visualization.

When the tech world is gravitating towards big data, GIS is gravitating away from it. At FOSS4G, the focus wasn't on big; the focus was very much on small. Notably, small data is defined as:

Small data is the term for a collection of data sets so small or niche that it becomes difficult to justify the cost and complexity of common database management tools or traditional data processing applications. The challenges include cost effective capture, storage, search, sharing, transfer, analysis, and visualization [6].

V. SMALL DATA AND IoT

The millions and billions of connected devices that make up the IoT generates a huge amount of data where the traditional data processing applications can no longer keep up.

Figure 2: Small data and IOT (courtesy [5])

Next, when it comes to Big Data, the challenges are picking and analysing the right data to take decisions, and make sure that the different kinds of data formats are able to communicate and exchange data with other device. But IoT is not only about big data, many IoT use cases only require small datasets of specific attributes of current state triggers to request the event like- the patient's blood sugar levels are high or not, the containers in the refrigerated kept at the optimal temperature or not, whether the soil have the right mixture of nutrients or not, the valve leaking or not.

Small data uses FOG computing unlike big data, where big data which uses cloud computing to process data. FOG computing is an architecture that depends on multiple devices and it allows data to be analysed prior to transmission so that instead of sending a massive amount of raw data, the information is analysed to produce small data which will be actionable insights helping to reduce the amount of bandwidth needed for the transfer [7].

Through these efficient business processes companies can save millions of dollars from the analysis of relatively small datasets. Small data knows what a tracked object is doing. Small data need not know why the tracked object is doing that.

Small data is used to determine current states and conditions or may be generated by analysing larger data sets.

IoT not only deals with big data, dealing adequately with small data is also as essential for making Internet of things a success.

Most businesses have woken up to the reality that data analysis is the lifeblood of a 21st century corporation. Yet they are not prepared for the incoming flood of data that these connected objects will provide. It's one thing to track your customers as they flow through your website, or observe that the yellow button gets way more clicks than the red button. It's another thing to track data created inside and outside a company about their employees, customers, and partners. We need new technologies to make sense of this glut of information. And this is going to change the way that big data companies like Hortonworks HDP 1.47%, Cloudera CLOUDERA 0.00%, IBM 1.55%, and others sell products and services [8]. Structure data can solve the above issues

That method also involves machine learning techniques that can help data analysis systems learn which inputs to prioritize, the same way you use context clues to pick out the handful of relevant emails in your inbox each morning against the hundreds of extraneous ones.
If the budding giants of big data—Hortonworks, Cloudera, and their ilk—don't adjust to meet this challenge, someone surely will. And even if the Internet of things arrives bit by bit, instead of as a tsunami, the data scientists, engineers, and developers who get this right over the next few years will be in prime position for both profits and influence over the development of our connected world [8].

Illustration of small data in case of IoT:

A wind turbine has a variety of sensors mounted on it to determine wind direction, velocity, temperature, vibration, and other relevant attributes. The turbine's blades can be programmed to automatically adjust to changing wind conditions based on the information instantly provided by small data.
Another example of small data is the use of smart labels on medicine bottles, which is used to determine where the medicine is located, its remaining shelf life, whether the seal of the bottle has been broken, and the current temperature conditions in an effort to prevent spoilage [9].

VI. SMALL DATA AND MARKETING

The main aspect of marketing is reliability. According to McKinney's, very few companies have achieved what we call "BIG IMPACT THROUGH BIG DATA". The information is so huge that it can't be tamed and managed as per expectations. We come across organizations take up large big data projects spending lot of resources, but the projects fail to take off or the results take long time to show up.

A global study sponsored by Teradata Corporation and conducted by, The Economist Intelligence Unit in 2014, reveals that CEOs often have an inflated view of their big data initiatives.

Here are some statistics of the global study:

- 47% of CEOs believe that all employees have access to the data they need while 27% of all respondents agree that they do.
- 43% of CEOs think relevant data are captured and made available in real time While 29% of all respondents agree they do.
- 38%CEOs are also more likely to think that employees extract relevant insights from data while 24% of all respondents and only 19% of senior vice presidents, and directors agree they do.
  From this global study [10] we conclude that big data has no value unless it offers meaningful insights to end users who can leverage the data and draw insights from it.

VII. INSTANCE OF SMALL DATA IN MARKETING

"Freshly introduced Hubspot Signals product, gives sales personnel immediate feedback to client activities such as opening newsletters, email queries or Facebook likes. In another small data example, Bonde pointed to the travel site Kayak, which can offer users an immediate estimate if the price of the trip, they are considering will increase or decrease based on the booking experience of other travellers [11].

The possibility of individuals owning their small data would mark a stunning change from the world of big data profiles and marketing.

At PeerJ, each article and each preprint has a counter for how many times that article or preprint has been viewed, downloaded and how many visitors have been on that page.

Previously, they were using a combination of Google Analytics (GA) to measure the page views and page visitors as well as custom process in varnish to detect pdf download. They found GA wasn't reliable enough and also return an incorrect value. So, their new process uses Amazon Web Services (AWS) Elastic Map Reduce (EMR) to process their log data and delivery individual article and preprint analytics.

They found that using EMR significantly reduced the learning curve around Hadoop and Hive. They were also impress with Hue's ability to immediately start to look at the data, which is great when you want to extract small datasets [12].

VIII. WHAT MAKES SMALL DATA DIFFERENT

Small data vs Big data

Case	BIG DATA	SMALL DATA
Data Sources	Data generated outside the enterprise from non-traditional data sources, Include: 1]Social media 2]Sensor data 3]Log data 4]Device data 5]Video 6]image [13].	Traditional enterprise data. Includes: 1]Enterprise Resource Planning transactional data 2]Customer Relationship Management (CRM) systems 3]Web Transactions

		4]Financial data
Volume	1]Terabytes 2]Petabytes 3]Exa bytes 4]Zettabytes	1]Gigabytes 2]Terabytes
Velocity	1]Often Real-time 2]Requires immediate response	1]Batch or near real-time 2]Does not always require immediate response
Variety	1]Structured 2]Unstructured 3]Multi-structured	1]Structured 2]Unstructured
Value	Complex, advanced, predictive business analysis and insights	1]Business intelligence, analysis and reporting
Cost	Expensive	Economic

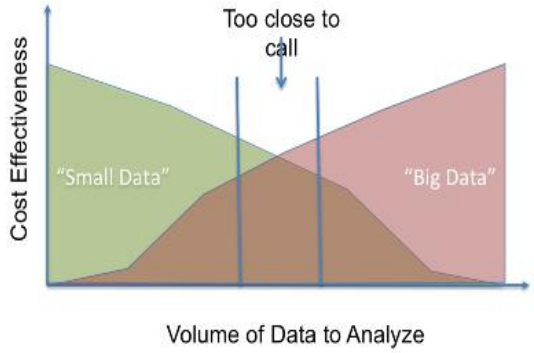

figure 1: Comparison –courtesy [5]

IX. TECHNICAL CHALLENGES OF SMALL DATA

The following explanation focuses on the technical challenges that is applicable to Small data includes:

1) Getting the Personal data APIs is highly challenging since one can Convincing service providers is a difficult task.
2) Test bed for prototype and learners
- Economy of scale in a shared testbed for rapid iterative exploration.
- Secure and private data handling, IRB, methods, tools.

3) Scalability/flexibility of database schema
- Track relationships among samples.
- Diverse context for new sample and data type.

4) Performance of search application.
5) Interoperability interfaces.
6) Data ingestion & quality control.

This challenges can be overcome by tools and frame works like and Hadoop can also be used for small data management [14].

X. CONCLUSION

Small data is useful in providing insights and it is actionable as well as understandable by the people. It allows organizations to exhibit by operating attainable data in small packets. For instance, small data can be used to provide information on how a particular machine part is performing at a given time. Now to find answers to 'why' it is performing in that way, we might need big data. So, going ahead we will see the 'big and small' complimenting each other. Small datasets are meaningful in a situation, for a particular purpose. Taking quality data over quantity of data is what makes sense. Return on data depends on our knowledge and judgment in collecting the right data and in preparing and using it well. Understanding the necessity of right data which is to be used will help staggeringly to take smarter business decisions based on the data"

It can be summed up that in this larger power of computing and growing interest in things like data mining, small data can be fascinated in industrial science. Despite, certain kinds of data depend on humans.

XI. REFERENCES

[1] The data science central, http://www.datasciencecentral.com/,-15th Jan 2016
[2] Definition of Small data, whatis.techtarget.com-24th Jan 2016
[3] ZDNET SPECIAL FEATURE: THE 21ST CENTURY DATA CENTER: http://www.zdnet.com/ -28th Jan 2016
[4] Martin Lindstrom," Small Data-The tiny clues that uncover huge trends"
[5] Small data vs big data, https://www.linkedin.com- 19th Feb 2016
[6] FORGET "BIG DATA", THE FUTURE OF GIS IS SMALL DATA, blog.mangomap.com-19th Feb 2016
[7] Why Small data is the future of IoT, ideacouture.com/iot-trends/why-small-data-is-the-future-of-iot/ -04th Mar 2016
[8] The Internet of Things Will Make Big Data Look Small by Tom Krazit, fortune.com-17th Mar 2016
[9] Forget Big Data – Small data is driving the internet of things by Mike Kavis, www.forbes.com,-22nd Mar 2016
[10] Start thinking 'small' data by Smita Vasudevan, www.dqindia.com/start-thinking-small-data/-25th Mar 2016

[11] 'Small Data' Analysis the Next Big Thing, Advocates Assert By Eric Lundquist, www.eweek.com,- 29th Mar 2016
[12] Peer blog-- Using big data tools for small data by PATRICK MCANDREW, peerj.com/blog/post/author/patrickpeerj-com/-30th Mar 2016
[13] Small data vs big data, https://datafloq.com/-1st April 2016
[14] P. Amuthabala, "OUTLOOK ON VARIOUS SCHEDULING APPROACHES IN HADOOP", International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397 Vol. 8 No.2 Feb 2016

AUTHOR PROFILE

SMITHA K.A – presently a student in the Department of Information Science & Engineering, JSS Academy of Technology, Bangalore, India.

SANGEETHA B – presently a student in the Department of Information Science & Engineering, JSS Academy of Technology, Bangalore, India.

Ramu B.S – presently a student in the Department of Information Science & Engineering, JSS Academy of Technology, Bangalore, India.

Dr. D.V.Ashoka- presently working as a Professor in the Department of Information Science & Engineering, JSS Academy of Technology, Bangalore, India. He got his PhD in Computer Science and Engineering from Dr. MGR University, Chennai, India, M.Tech in Computer Science and Engineering from VTU, Belgaum, India and B.E. in Computer Science and Engineering from Kuvempu University, India. He is an approved research guide for many Indian universities and has over 20 years of experience in Teaching, Research and Administration. He has professional membership with FELLOW IEI, IEEE, MISTE, MCSI and MIAENG. He is one of the National Award winners "Rashtriya Ekta Samman-2013".His research interests are in Operating System Virtualization, Knowledge Engineering, Communication, network security and Data Mining.